IN SILICO CHARACTERIZATION AND EXPRESSION PROFILING OF THE DIACYLGLYCEROL ACYLTRANSFERASE GENE FAMILY (DGAT1, DGAT2, DGAT3 and WS/DGAT) FROM OIL PALM, *ELAEIS GUINEENSIS*.

Rozana Rosli[1,2], Pek-Lan Chan[2], Kuang-Lim Chan[2], Nadzirah Amiruddin[2], Eng-Ti Leslie Low [2], Rajinder Singh[2], John L Harwood[3], Denis J Murphy[1]*

[1] Genomics and Computational Biology Research Group, University of South Wales, Pontypridd, CF37 1DL, United Kingdom

[2] Advanced Biotechnology and Breeding Centre, Malaysian Palm Oil Board, No 6, Persiaran Institusi, Bandar Baru Bangi, 43000 Kajang, Selangor, Malaysia

[3] School of Biosciences, University of Cardiff, Cardiff, CF10 3AX, United Kingdom

*Corresponding author

E-mail: denis.murphy@southwales.ac.uk

Highlights

- Genes encoding four distinct functional families of diacylglycerol acyltransferases (DGAT) enzymes were characterised in the genome of the African oil palm, *Elaeis guineensis* and in 12 other oil crop or model/related plants.
- The oil palm genome contains respectively three, two, two and two distinctly expressed functional copies of the DGAT1, DGAT2, DGAT3 and WS/DGAT genes.
- Comparison of the predicted DGAT sequences was consistent with the *E. guineensis* DGAT1 being ER located with its active site facing the lumen while DGAT2, although also ER located, had a predicted cytosol-facing active site.
- In contrast, DGAT3 and WS/DGAT in *E. guineensis* are predicted to be soluble, cytosolic enzymes.
- Evaluation of *E. guineensis* DGAT gene expression in different tissues and developmental stages suggests that the four DGAT groups have distinctive physiological roles and are particularly prominent in developmental processes relating to reproduction, such as flowering, and in fruit/seed formation especially in the mesocarp and endosperm tissues.

Abstract

The diacylglycerol acyltransferases (DGAT) (diacylglycerol:acyl-CoA acyltransferase, EC 2.3.1.20) are a key group of enzymes that catalyse the final and usually the most important rate-limiting step of triacylglycerol biosynthesis in plants and other organisms. Genes encoding four distinct functional families of DGAT enzymes have been characterised in the genome of the African oil palm, *Elaeis guineensis*. The contrasting features of the various isoforms within the four families of DGAT genes, namely DGAT1, DGAT2, DGAT3 and WS/DGAT are presented both in the oil palm itself and, for comparative purposes, in 12 other oil crop or model/related plants, namely *Arabidopsis thaliana*, *Brachypodium distachyon*, *Brassica napus*, *Elaeis oleifera*, *Glycine max*, *Gossypium hirsutum*, *Helianthus annuus*, *Musa acuminata*, *Oryza sativa*, *Phoenix dactylifera*, *Sorghum bicolor*, and *Zea mays*. The oil palm genome contains respectively three, two, two and two distinctly expressed functional copies of the DGAT1, DGAT2, DGAT3 and WS/DGAT genes. Phylogenetic analyses of the four DGAT families showed that the *E. guineensis* genes tend to cluster with sequences from *P. dactylifera* and *M. acuminata* rather than with other members of the Commelinid monocots group, such as the Poales which include the major cereal crops such as rice and maize. Comparison of the predicted DGAT protein sequences with other animal and plant DGATs was consistent with the *E. guineensis* DGAT1 being ER located with its active site facing the lumen while DGAT2, although also ER located, had a predicted cytosol-facing active site. In contrast, DGAT3 and some (but not all) WS/DGAT in *E. guineensis* are predicted to be soluble, cytosolic enzymes. Evaluation of *E. guineensis* DGAT gene expression in different tissues and developmental stages suggests that the four DGAT groups have distinctive physiological roles and are particularly prominent in developmental processes relating to reproduction, such as flowering, and in fruit/seed formation especially in the mesocarp and endosperm tissues.

Key words: diacylglycerol acyltransferase (DGAT); endosperm; kernel; mesocarp; oil palm (*Elaeis guineensis*); triacylglycerol

Introduction

The African oil palm, *Elaeis guineensis*, is the most important global vegetable oil crop in terms of both yield efficiency (tonnes oil/hectare) and overall volume of production [1–3]. Oil palm fruits contain two types of storage oil located respectively in the fleshy mesocarp tissue of the fruit and the triploid endosperm tissue of the seed kernel. In contrast to many prominent temperate oilseed crops, such as soybean and rapeseed, where the storage oil accumulates in the embryo, oil storage in palm fruits occurs in non-embryo, maternally-derived tissues and may therefore be subject to different forms of genetic regulation. Moreover, in contrast to the storage role of the seed/kernel oil, the major role of the mesocarp oil is as an attractant to potential animal vectors that serve to disseminate ingested seeds. Again this distinctive role may result in different evolutionary constraints that could affect genetic regulation of mesocarp versus kernel oils.

The mesocarp oil, commonly referred to as palm oil (PO), is mostly made up of triacylglycerols (TAGs) containing long-chain C16 and C18 fatty acids (about 44% palmitate, 39% oleate and 11% linoleate) [4]. In contrast, the seed endosperm oil commonly referred to

as palm kernel oil (PKO), is enriched in medium-chain C12 and C14 fatty acids (about 48% laurate and 16% myristate). Together these two palm-derived oils account for about 38% of global production of commercial vegetable oils [1]. Further information about palm oil has been reviewed by Sambanthamurthi et al. (2000) and Murphy (2014). Two key recent developments in oil palm research have been the publication of the genome sequence in 2013 [6] and the compilation of an updated gene model dataset for the species in 2017 [7].

Like the vast majority of plant storage oils, the two types of palm oil are overwhelmingly made up of TAGs that are synthesised in the ER and then accumulated as cytosolic lipid droplets (LDs) [8]. The overall nature of the pathways involved in TAG biosynthesis in plants is well established [9–12], although several new enzymes have recently been discovered [10] and the relative contributions of different reactions and the details of their regulation have yet to be fully resolved [13]. Central to lipid assembly in oil crops is the Kennedy pathway [14] which converts glycerol 3-phosphate in four steps to TAG using acyl-CoAs as the source of fatty acyl residues. The final reaction in this sequence is catalysed by diacylglycerol acyltransferase (DGAT) (diacylglycerol:acyl-CoA acyltransferase, EC 2.3.1.20).

Numerous studies have pointed to the DGAT reaction being critical for TAG assembly and in several cases it has been shown to limit carbon flux from lipid precursors towards TAG accumulation [12]. Thus, in *Brassica napus*, the DGAT substrate, diacylglycerol (DAG), accumulates during periods of rapid lipid formation [15,16]. Moreover, DGAT had the lowest activity (as measured *in vitro*) in extracts of developing seeds and DAG levels were the highest of all the Kennedy pathway intermediates in seed tissues [17]. In addition, an *Arabidopsis thaliana* mutant (ASI1) with reduced DGAT activity, had a decreased TAG/DAG ratio compared to wild type plants [18] while Zou et al. (1999) showed that this phenotype was due to a mutant allele of the DGAT1 gene [20]. Furthermore, seed-specific over-expression of a DGAT1 gene led to increased oil content in transgenic plants [21]. In *B. napus* cv. Westar, the Kennedy pathway and associated reactions exerted stronger control over carbon flux to TAG than did fatty acid provision [22]. Also, the overexpression of DGAT1 resulted in lower flux control values for overall TAG assembly [23]. In addition, transgenic *B. napus* plants with an enhanced DGAT1 activity exhibited increased oil accumulation in field trials [24], again emphasising the importance of DGAT gene expression and enzyme activity for overall oil yields at the level of the crop.

Similar studies with other oil-accumulating plants, such as *Cuphea*, lupin, soybean and *Linum* species, have supported the notion that DGAT activity is important in determining the overall levels of TAG accumulation [12,25]. In general, the rise in DGAT activity during TAG accumulation parallels that of other Kennedy pathway enzymes [26]. In contrast, activities of enzymes involved in *de novo* fatty acid biosynthesis do not show such good correlations with oil accumulation [27–29] This suggests that TAG assembly is more tightly controlled than fatty acid synthesis [12]. In the case of *E. guineensis*, the regulation of TAG synthesis has been studied in detail using callus cultures. The data from control analysis experiments showed that flux control is shared between fatty acid synthesis and TAG assembly [30,31]. In addition, the contribution of DGAT to TAG assembly was assessed directly by inhibition experiments *in vitro* [32]. Further information about the use of control analysis to give quantitative information about lipid biosynthesis in *E. guineensis* has been described by Ramli et al. (2009).

The importance of DGAT in contributing to the regulation of oil accumulation in crops has led to its use, not only in single-gene, over-expressing transgenic lines [21,23,34], but also in plants manipulated for both 'push' and 'pull' activities. In the former, carbon supply for lipid synthesis is increased (push) while, in the latter, DGAT activity, as the final step in TAG formation, is raised (pull). For example, overexpression of the transcription factor WRI1 (WRINKLED1) and of DGAT1 in tobacco seeds led to enhanced TAG accumulation compared to that expected by an additive effect [35]. Furthermore, a combination of DGAT expression and PGM (phosphoglucomutase) suppression has been used as an example of a combined push/pull strategy to boost TAG yields in the important oil crop, soybean [36,37]. In addition, a combination of DGAT and LEC2 (LEAFY COTYLEDON 2) gene overexpression has been used to increase TAG accumulation in tobacco [38]. The concept of using other enzymes in addition to DGAT in order to raise oil yields has also been used to increase TAG accumulation in tobacco leaves, which do not normally accumulate high levels of TAGs [39]. In this innovative study, carbon flux was increased through both fatty acid synthesis ('push') and TAG formation ('pull'), while at the same time TAG-rich lipid droplets were stabilised via oleosin over-expression ('package') and minimising further metabolism by silencing the SDP-1 lipase ('protect'). This led to an incremental, step-wise increase in the ectopic accumulation of TAG to the remarkably high levels of >30% of leaf dry weight [39].

Experimentally measured DGAT activity was first reported by Weiss et al. (1960) and several different types of DGAT enzyme have since been described in plants [12,41,42]. As recently as 2011 there appeared to be just two DGAT enzymes both in plants and in other eukaryotes, namely DGAT1 and DGAT2 [43]. Both DGAT1 and DGAT2 are membrane-bound (normally on the ER) enzymes but they are otherwise structurally very distinct from each other. It therefore seems likely that these two enzymes originally evolved separately but became functionally convergent as they acquired similar types acyltransferase activity involving DAG substrates [43], albeit possibly with different roles in ER-based TAG formation in different plant tissues. More recently, a third putative DGAT isoform, a soluble enzyme termed DGAT3, was discovered and there are preliminary reports that this enzyme has DGAT activity and may also participate in a cytosolic pathway of TAG biosynthesis [25,44,45]. Finally, a fourth DGAT activity, a bi-functional DGAT/wax ester synthase (WS/DGAT) has been described in a wide range of organisms from bacteria to plants [46]. The primary function of WS/DGAT is believed to be the formation of surface wax esters, although it has been suggested that this enzyme is also responsible for making small amounts of TAG [25,46,47]. Interestingly WS/DGAT is a very diverse protein family with some members shown to be soluble in cells while others are membrane-bound [48–50]. As with the DGAT1 and DGAT2 genes, both DGAT3 and WS/DGAT have very distinct evolutionary pathways and appear to have originated independently of each other [25]. In all of the land plant genomes and at least one algal genome analysed to date some or all the four DGAT gene families have multiple copies, implying that the duplication events responsible for this probably occurred prior to Streptophyte diversification [25,51].

Genetically speaking, oil accumulation in plant tissues is a complex quantitative trait that involves numerous genes. Efforts to increase oil yields in commercially valuable crops, such as *E. guineensis* and oilseeds like soybean, require the identification of the specific genes that regulate this highly desirable agronomic trait so that breeders can focus on variation involving those key genes [52] in a similar manner to efforts to manipulate the acyl quality of the oil [53]. It is becoming increasingly apparent that DGAT activity is pivotal to increasing

oil yield in seeds, as demonstrated by the significant increases in TAG accumulation when DGAT genes are over-expressed in transgenic plants [12,25,54]. *E. guineensis*, as the world's most productive edible (and industrial) oil crop, is an important contributor to global food security with several studies indicating that demand for the oil will continue to increase substantially over the coming decades [2].

This has led to concerns that the forecast increased demand for palm oil might lead to further conversion of sensitive tropical habitats to *E. guineensis* plantations [2]. However, an alternative strategy would be to increase oil yields in *E. guineensis* fruits themselves so that more oil can be produced from exisiting smallholder and commercial plantations. This would reduce the requirement for the conversion of additional land for *E. guineensis* cultivation [2]. A further challenge is the improvement of oil quality in order to expand markets for palm oil, e.g. by reducing the saturate content and increasing oleic acid levels in order to compete more effectively with premium edible vegetable oils such as olive and sunflower oils and also to reduce free fatty acid levels by inhibiting or removing lipase gene expression in freshly picked palm fruits [55].

In order to fulfil the strategy of improving palm oil yield and quality, it is necessary to improve our understanding of the regulation of TAG accumulation and especially the role of DGAT in the non-seed tissues of *E. guineensis* fruits, namely the mesocarp, where relatively few studies have been performed to date [56–58]. In this study, we describe the genomic architecture of the various isoforms within the four classes of DGAT genes in *E. guineensis*, namely DGAT1, DGAT2, DGAT3 and WS/DGAT as compared with 12 other plant species. We have also evaluated DGAT gene expression in a range of tissues, including mesocarp, seed kernel, and selected vegetative tissues, as well as at different developmental stages.

Methods

Identification of Diacylglycerol Acyltransferase genes

Publicly available sequences from several databases were downloaded for 13 plant species, namely *A. thaliana*, *B. distachyon*, *B. napus*, *E. guineensis*, *E. oleifera*, *G. max*, *G. hirsutum*, *H. annuus*, *M. acuminata*, *O. sativa*, *P. dactylifera*, *S. bicolour* and *Z. mays*. Ortholog analysis was performed using Orthomcl2.0 [59] using default parameters. A list of curated DGAT genes from Arabidopsis was used as a reference to identify DGAT orthologs from the Orthomcl data (Supplementary S1 file). To search for other *E. guineensis* DGAT genes especially those which are singletons, a hidden Markov model (HMM) profile was built from orthologous amino acid sequences and used as a query for hmmsearch from the HMMER3 package [60] against our *E. guineensis* gene models [7]. Sequence similarity searches, using BLASTP program [61] against reference sequence protein database (Viridiplantae), were performed to identify additional DGAT genes in *E. guineensis*.

Analysis of derived amino acid sequences and protein domains

Protein domains were identified using InterPro (http://www.ebi.ac.uk/interpro/), PfamScan (http://pfam.xfam.org/search#tabview=tab1) and NCBI CDD (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Motif analyses by selection of up to 20 motifs were performed using Motif Multiple En for Motif Elicitation (MEME) [62]. Transmembrane domains were predicted by using TM domains plugins Transmembrane Prediction Tool version 0.9 Geneious version 10.2.3 (http://www.geneious.com, Kearse et al. 2012) for the classification of DGAT family members. The lengths of the TMs is 21 - 22 residues. Multiple sequence

alignments of the cluster members were performed by MUSCLE [64] and the sequences were compared and visualized using Bioedit tools [65]. Protein relationship trees were built using Molecular Evolutionary Genetics Analysis (MEGA 7) [66] with Neighbour-joining (NJ) methods [67] using Poisson correction method [68] and Maximum Likelihood method based on JTT matrix-based model [69] .

Localization of DGAT genes in the E. guineensis genome and gene structure analysis

TBlastN was used to identify the DGAT scaffold sequence from our genome assembly data. The Exonerate program (version 2.2.0) [70] with protein2genome model parameters was used to find the number of exons and introns and to predict the location of DGAT gene in our EG5.1 Genome build [7]. Gene structures between the detected orthologs were generated using WebScipio interface version Scipio v1.5 [71,72]. The alignments of gene structures were generated and compared with GenePainter [73].

Expression profiles of DGAT genes in endosperm, mesocarp and vegetative tissues

RNA seq data from endosperm, mesocarp and vegetative tissues were read mapped to the *E. guineensis* genome P5-build using the Tuxedo suite [74] and linked to our *E. guineensis* gene model [7]. Sequencing of transcriptome libraries was performed using Roche/454 and Illumina HiSeq 2000 machines. The complete data sets are publicly available from NCBI BioProjects PRJNA201497 and PRJNA245226. For Illumina HiSeq 2000 data, the $\log_{10}$ FPKM expression profiles of four DGAT gene families were visualized using line graphs while heatmaps were generated as outputs for the Roche/454 data.

Results and Discussion

DGAT gene copies and phylogenetic analysis

Phylogenetic and amino acid motif analyses corresponding to the four DGAT families as retrieved from the genomes of *E. guineensis* and other plants of interest are shown in Figs. 1, 2, 3 and 4. Note that we also used two independent phylogenetic methods, namely maximum likelihood and maximum parsimony, to confirm the phylogenies shown here (see Additional file 1. In the case of DGAT1 (see Fig. 1A) we retrieved three apparently full length sequences, one of which (EgDGAT1_2) clustered with an ortholog from date palm (PDK30s995161g003) while the other two (EgDGAT1_1 and EgDGAT1_3) formed an adjacent but distinct cluster with EgDGAT1_1 additionally forming a sub-branch with nine very similar isoforms from *E. oleifera*. All of these oil and date palm sequences were in a larger cluster that additionally included two sequences from banana, *M. acuminata*. Banana and date palm and *E. guineensis* and *E. oleifera* are all members of the Commelinid monocots group. The Commelinids are an apparently monophyletic taxon that includes many commercially important species such as the palms, bananas, ginger and all members of the grass family including the major cereal crops such as rice, wheat and maize plus the model species *B. distachyon*. However, more recent molecular evidence has questioned the precise place of palms versus grasses in the Commelinid monocots group, not least because while the grasses have some of the most rapid recorded rates of molecular change, the palms have some of the slowest [75–78]. Interestingly, in Fig 1A the palm and banana sequences clearly cluster in a distinct group that is separated from the other members of the Commelinids, which are distributed within a large group that includes several non-monocot species such as soybean, *G. max*, and sunflower, *H. annuus*.

In the case of DGAT2 the *E. guineensis* genome contained two apparently full length sequences (EgDGAT2_1 and EgDGAT2_2) that formed a distinct cluster with putative banana and date palm orthologs (see Fig. 2A). EgDGAT2_1 was in a sub-branch with three sequences from *E. oleifera*, banana and date palm, while EgDGAT2_2 formed a separate sub-branch with sequences from *E. oleifera* and banana only. All seven of the palm/banana DGAT2 sequences formed a discrete cluster that was adjacent to but distinct from a cluster of Poales sequences that included maize and rice.

Compared to DGAT1 and DGAT2, the DGAT3 genes are a much more recently discovered family that encodes soluble proteins in plants [25,44,45]. A recent comparative study showed that there were either one or two DGAT3 genes in the genomes of a wide range of plants ranging from simple marine algae such as *Volvox carteri* and *Ostreococcus lucimarinus* to major crop species such as soybean and rice [25]. The *E. guineensis* genome contained two clearly distinct DGAT3-like sequences, one of which (EgDGAT3_2) formed a small cluster with two isoforms from *E. oleifera* and one date palm sequence, while the other gene (EgDGAT3_1) clustered with a single *E. oleifera* sequence (see Fig. 3A). All of these sequences formed a cluster with two sequences from banana and clearly constituted a different grouping from other plant DGAT3 sequences, including those from other monocots such as rice and maize plus dicots such as soybean and rapeseed.

The WS/DGAT gene family varies considerably in the number of putative, annotated gene sequences in different plant species with maize and soybean apparently only having one gene copy while cotton has five and *A thaliana* has no fewer than 11 gene copies [25,47]. The *E. guineensis* genome contained two clearly distinctive WS/DGAT-like sequences, with EgWS/DGAT_1 clustering with two isoforms from *E. oleifera* and one from date palm, while EgWS/DGAT_2 clustered in a completely different branch with another *E. oleifera* isoform and date palm sequences (see Fig. 4A). Interestingly, 12 WS/DGAT sequences from soyabean cluster together with the EgWS/DGAT_2 sequence. A third sequence, labelled as EgWS/DGAT_3 in Fig 4A, was rejected as a potential functional WS/DGAT as only half of the full length sequence that was present in the other proteins was present in this putative WS/DGAT and we conclude that the *E. guineensis* genome contains just two WS/DGAT genes. Unlike the other three DGAT gene families, the putative WS/DGAT genes shown in Fig 4A were not as discretely clustered into palm/banana, Poales and dicot groups and, as discussed below, the derived WS/DGAT amino acid sequences of *E. guineensis* were also more divergent from each other.

Structural and motif analysis of DGAT-like sequences in E. guineensis and other plants

Motif analysis using MEME was used to characterise the derived amino acid sequence domains in all four DGAT gene families in *E. guineensis* and the other 12 analysed plant species as shown in Figs. 1B to 4B. Within the 13 plant species that we surveyed, we identified ortholog sequences as follows: a total of 39 DGAT1 sequences (21 genes, 18 isoforms), 39 DGAT2 sequences (26 genes, 13 isoforms), 19 DGAT3 sequences (17 genes, 2 isoforms) and 97 WS/DGAT sequences (82 genes, 15 isoforms) (Supplementary S1 file). In terms of the enzymatic classification, DGAT1 is a member of a eukaryotic ER-located 50-60

kDa protein family annotated as 'Diacylglycerol O-acyltransferase 1' in UniProt. DGAT2 belongs to a quite distinct 30-40 kDa eukaryotic ER-located protein family annotated as 'Diacylglycerol O-acyltransferase 2' in UniProt. DGAT3 is soluble 40-50 kDa protein family annotated as 'Diacylglycerol O-acyltransferase 3, cytosolic' in UniProt and finally WS/DGAT is classified as 'similar to the O-acyltransferase WSD1 group'. For the WS/DGAT sequences, the locations of key individual residues such as histidines and longer conserved motifs shown in Fig 4B and S2 are similar to those found in other well characterised members of the WS/DGAT family, also referred to as WSD in some reports [49,50]. In particular the putative active site residues HHxxxDG are absolutely conserved as shown in Supplementary Fig. S2D at positions 239-245. We are therefore confident that the members of all four DGAT families have been identified in the 13 species that were analysed.

Fig. 1B shows the conserved motifs structures that were identified by MEME analysis in EgDGAT1. However, the three motifs LHSAAAAVVVQ, EDSSKTSLPSAEDSNTDSGEDSGVDTSSDADTRDRVVDGVDREE, GEEKAG and NGEKYDDGAGRAEGQEAGVGV were missing in EgDGAT1_2. The absence of motif KGDMSSSCJEIDNMKGPSFKSLVYFMVAP in EgDGAT1_1 also differentiates it from EgDGAT1_3 and this motif is also not present in EgDGAT1_2. Although the EgDGAT1 and EgDGAT2 sequences are broadly similar to the corresponding genes in date palm, a more detailed motif analysis revealed that the date palm protein sequences were missing several motifs. For example in DGAT1, PDK_30s995161g003 is missing five motifs towards the C terminal part of the sequence while in the case of DGAT2, the date palm sequence PDK_30s855921g004 is missing almost half of the N terminal part of the sequence (Fig. 2B).

The data also indicate that there are two types of DGAT3 in *E. guineensis*. One of the DGAT3 sequences (EgDGAT3_2) was very similar to the date palm PDK_30s923651g005 in containing the conserved motif EENNYALKLGPECSNTSATTSSSDSCGCCSNSIPVV DRPMD, which is absent from EgDGAT3_1 and from two DGAT3 sequences in banana, namely GSMUA_Achr2P07940_001 and GSMUA_Achr1P25080_001 (Fig 3B). In Fig 4B, motif analysis enabled the identification of two distinct types of *E. guineensis* WS/DGAT. The first, EgWS/DGAT_1, was found to be common to all 12 plant species analysed while the other sequence, EgWS/DGAT_2, was only found in six of the plant species, namely *E. oleifera*, *G. hirsutum*, *G. max*, H. *annuus*, *M. acuminata*, *P. dactylifera.* A third *E. guineensis* WS/DGAT sequence, shown in Fig 4B as EgWS/DGAT_3, was almost identical to the N-terminal half of EgWS/DGAT_2 but the C-terminal half of the sequence was missing and this may be a truncated pseudogene or a sequencing error.

Our data from the *E. guineensis* genome and the other comparative plant genomes, as generated using Geneious tools and additional plug-ins, show that the predicted amino acid sequences of the four major DGAT protein classes differ considerably in both composition and their overall length. In particular, the two major types of plant DGATs, both of which are predicted to be membrane bound, namely DGAT1 and DGAT2, shared very little amino acid sequence similarity [79,80]. While DGAT1 ranged from about 500 - 540 amino acids (aa), DGAT2 was much shorter at about 330 aa. Two types of DGAT1 were characterized when we compared the three annotated sequences from the latest *E. guineensis* gene model [7] with those described in [81]. Using tblastn searches it was found that EgDGAT1_2 [7] is similar to EgDGAT1-1 [81], while EgDGAT1_3 [7] were similar to EgDGAT1-2 [81]. However, EgDGAT1_1 only showed 63% identity (339/535; match length 516/516 aa) with EgDGAT1-1 for the first hit and 82% (293/357; match length 331/516 aa) for the secondary hit to EgDGAT1-2. Multiple sequence alignments (Supplementary Fig. S2A) show an

important conserved sequence X**H**XXX(X)**D** motif box 1 [82] and putative functional motifs were predicted for four EgDGAT by using Prosite database [83] (Additional file 2). It has been shown that the His (H) and Asp (D) residues can be replaced with Arg (R) and Glu (E) residues in box 1 to give the same function in the cases of plant, mouse and human ortholog sequences [21].

In all of the analysed plant genomes the predicted DGAT2 proteins were over 200 aa shorter than DGAT1 at ~330 aa. There were two *E. guineensis* DGAT2 sequences, EgDGAT2_1 and EgDGAT2_2 and these were 99% and 55% identical to the previously reported sequences from [81], respectively. In parallel, the EgDGAT2 transcript (XM_010933834), which was isolated from mesocarp tissues, was compared with EgDGAT2_1 and this indicated that the start codon for EgDGAT2_1 might be located after 35bp. In Fig. 2B, EgDGAT2_2 is missing most of a conserved motif located towards the C terminus. Similarity search against RefSeq showed 52% similarity with *A. thaliana* AT3G51520 and 55% with *G. max* DGAT2D (LOC100784657). Sequence conservation is shown in alignment of 33 orthologous sequences in Supplementary Fig. S2B.

Apart from the ubiquitous occurrence of DGAT1 and DGAT2 in plants, two other DGAT-related genes have been described, namely the soluble DGAT3 and the soluble multifunctional WS/DGAT [25,44,45]. The putative DGAT3 orthologs from *A. thaliana*, (At1g48300) and from *A. hypogaea* [44], were compared with the two *E. guineensis* DGAT3 sequences. Similar to DGAT2, DGAT3 proteins have predicted lengths of 340 - 360 aa but their amino acid sequences are otherwise very different (Fig. 3B). Of the two *E. guineensis* DGAT3 sequences, EgDGAT3_1 and EgDGAT3_2 had respectively 36% and 39% amino acid identical to *A. thaliana* At1g48300. Based on the sequence alignment data, EgDGAT3_2 had 33% identity and an overall 339 aa match length with the predicted peanut DGAT3 (AAX62735.1) although the alignment contained relatively high portion of gaps (21%). The data in Supplementary Fig. S2C demonstrate the uniqueness of DGAT3 sequences in all species. As reported by [25], DGAT3 was found be present as either one or two gene copies in all of their analysed plant species. From Orthomcl analysis, we deduced that there was only one copy of the DGAT3 gene in grasses such as rice, maize and Brachypodium, but two gene copies in *E. guineensis* and banana as well as in the selected dicots including soybean, cotton and rapeseed (Fig. 3B).

In plants, a bi-functional DGAT/wax ester synthase (WS/DGAT) was identified in Arabidopsis by Li et al. (2008). Although the primary function of this enzyme is believed to be the formation of surface wax esters, it has been suggested that it is also responsible for making small amounts of TAG [25]. In our ortholog analysis, it is interesting to note that the sequence of EgWS/DGAT_1 was relatively dissimilar to EgWS/DGAT_2. In view of the doubts about the functionality of the EgWS/DGAT_3 sequence (see above), similarity searches using blastp against Reference proteins database (refseq_protein) were performed to further validate this sequence. The results showed that sequence is 99% identical to the N-terminal half of Genbank accession XP_010935670 (EgWS/DGAT_2). Several analyses were then performed for both sequences to verify these results and the data suggest that EgWS/DGAT_3 is not a full-length functional DGAT3 as it is missing key motifs and that an update is required of the P5 gene model annotation [7]. For a full alignment of orthologs of the WS/DGAT family, see Supplementary Fig. S2D.

Topology and active site analysis

We used Geneious version 10.2.3 to estimate the number of putative transmembrane domains (TM) in DGAT proteins in all 13 analysed plant species as shown in Supplementary Fig. S3A. Note that we also ran several other algorithms such as TMpred, TopPred and TMHMM with varying degrees of consistency, but overall the Geneious package was found to give the most consistent predictions and gave similar results to those found previously with non-plant DGAT sequences, as discussed below. We found that for each DGAT family the number of TM can vary considerably even within a single species. There are two major likely explanations for this variation. Firstly the original genomic data and gene models/annotations upon which the derived amino acid sequences are based can vary in quality and reliability between the different plant species. This may give rise to partially erroneous or incomplete sequences that affect the predicted TM score. A second possibility is that in a given plant genome there may be several similar copies of a particular DGAT gene but that these have subsequently diverged structurally and/or functionally, e.g. so that they now have different substrate preferences and/or subcellular locations. In such cases the number of TM domains might differ but the proteins remain members of the particular DGAT family.

For DGAT1, the number of TM domains in all of the plant species varied between four and ten (Supplementary Fig. S3B). However, the most common pattern was for a cluster of four TM towards the N-terminus, with one or two in the centre, and three close to the C-terminus. The three *E. guineensis* DGAT1 isoforms followed this most common distribution in having eight putative TM domains with a 4-1-3 distribution at similar locations in the protein sequence. For an ER-located protein, this distribution of TM domains would give rise to a molecule with a large (100- to 150-residue) cytosolic N-terminal domain, plus two smaller (50- to 100-residue) centrally located cytosolic and ER luminal domains. This is similar to the predicted topological orientation of mammalian DGAT1 proteins, which also have a putative C-terminal proximate ER-luminal domain as well as 7-9 TM domains [42,84]. Interestingly, although hydropathy plots of the murine DGAT1 protein strongly predicted eight TM domains, topological studies were consistent with only three TM domains, although the end result was still a protein with a cytosolic N-terminal and a large C-terminal domain that included the putative active site as discussed below [85]. As with mammalian DGAT1 proteins, the three *E. guineensis* DGAT1 isoforms each contain a putative C-terminal proximate ER-luminal loop domain that includes the probable binding sites for the two enzyme substrates. These are (i) an acyl binding site, which includes the motif [FYxDWWN] that is highly conserved between DGATs and acyl-CoA cholesterol acyltransferase enzymes (in *E. guineensis* FGDREFYRDWWNAKT) and (ii) the putative diacylglycerol binding motif [HKWCIRHFYKP] that is also found in protein kinases C and diacylglycerol kinases (in *E. guineensis* NMPVHRWNIRHVY) [84]. In *E. guineensis* DGAT1 these two substrate binding domains are located directly adjacent to one another in the region of residues 445 to 480, which is predicted to be part of a larger non-cytosolic, i.e. ER luminal, domain as described above and depicted graphically in Supplementary Fig. S4. This is consistent with experimental studies showing that DGAT1 and DGAT2 are ER-bound membrane proteins in both plants and animals [79,86] and that DGAT1 from rapeseed has a cytosol-facing N-terminal region [87].

Although DGAT2 is also a DAG-active acyltransferase, like DGAT1, it is part of a quite separate, evolutionarily conserved, gene family that is highly expressed in tissues that synthesise and accumulate large amounts of TAG. In animals this includes adipose tissue, liver, small intestine, and mammary gland while in plants it mainly includes embryo and endosperm of oil-rich seeds and mesocarp of oil-rich fruits [86,88]. The wider family of

acyltransferases to which DGAT2 belongs includes monoacylglycerol acyltransferases (MGAT) 1–3 and wax synthases 1 and 2, all of which contain a highly conserved four-amino-acid sequence, HPHG, that is part of the larger so-called MBOAT motif of this protein superfamily [42,86,89]. Mutagenesis of amino acids within this sequence, and in particular the two histidine residues, markedly impaired the catalytic function of DGAT2 and suggested that this region is part of the active site [90]. DGAT2 sequences in animals and plants also contains several other highly conserved regions but their functional importance has yet to be determined [42]. Our analysis predicted two TM domains in EgDGAT2_1 and three in EgDGAT2_2, although the N-terminal proximate TM domain in EgDGAT2_2 was less confidently supported so both isoforms may contain just two TM domains (Supplementary Fig. S3B). This is similar to another report that the *E. guineensis* DGAT2 labelled as XM_010933834 contains two transmembrane domains [57] and is consistent with data from other organisms [42]. In nearly all cases the plant DGAT2 sequences that we analysed were predicted to have a short cytosolic N terminal domain followed by two TM domains and a large (>250 aa) cytosolic C terminal domain that included the conserved HPHG active site motif, which is similar to the topology found in murine DGAT2 [90].

These analyses, both here and elsewhere, of the ER-bound DGAT1 and DGAT2 proteins are consistent with the predicted topological structures shown in Supplementary Fig. S4 where DGAT1 has a luminal active site domain whereas in DGAT2 the active site is cytosolic, as also discussed by [86]. These different topologies are probably significant for the biological roles of DGAT1 and DGAT2. Hence it has been suggested that DGAT2 with its cytosol-facing active site is principally involved in bulk TAG formation as cytosolic lipid droplets – the so-called 'overt' activity – while DGAT1 has a different role, possibly in recycling exogenous TAG via the so-called 'latent' luminal activity as found in animals [86,91–93]. This is also consistent with the finding that recombinant tung tree DGAT1 and DGAT2 proteins expressed in tobacco cell lines localised to distinct punctate regions of the ER and are therefore likely to be located in different ER subdomains [79]. In contrast to DGAT1 and DGAT2, our analysis of *E. guineensis* and other plant DGAT3 and WS/DGAT gene products did not provide convincing evidence for the presence of TM domain sequences in most cases (Supplementary Figs. S3A and S3B). Only two out of 19 analysed DGAT3 sequences contained a single putative TM region located close to the C-terminus and well away from the predicted active site region. The situation for WS/DGAT was more complex, which is not surprising given the highly variable amino acid sequence compositions of this protein family in both plants and animals. Out of 97 plant sequences that we analysed, 52 were predicted to be cytosolic proteins with no TM domains while 31 were predicted to have one TM domain and 14 were predicted to have two TM domains (Supplementary Fig. S3A). Interestingly, however, regardless of whether each WS/DGAT has 0, 1 or 2 TM domains, the protein was still predicted to have a large cytosolic N-terminal region that included a highly conserved putative active site motif, HHXLGDG (Supplementary Fig. S2D). Therefore our data are consistent with previous reports that DGAT3 and WS/DGAT are soluble, cytosol-facing or cytosol-located proteins [25,44,45].

Gene Structure Analysis

The number and locations of introns and exons in the four DGAT gene families were predicted using Webscipio (Supplementary S1 file) and (Supplementary Fig. S5). By mapping the gene structure and aligning protein sequences of the closest related orthologs in the same clade we can derive additional useful information on conserved regions in the genes

and their encoded proteins. Fig. 5 shows the result of mapping protein alignments to gene structures in the four different DGAT groups. DGAT1 is a relatively long gene sequence and has 20 predicted exons which are relatively well conserved between *E. guineensis* and its close relatives. Hences, comparisons between the DGAT1 families show that EgDGAT1_2 is similar to an ortholog in date palm while EgDGAT1_1 and EgDGAT1_3 is more related to one in banana. There is a high possibility that DGAT1 from date palm is not full length as there is a missing of exon at 3' end of the sequence. The major difference between the two types/copies of EgDGAT2 is the presence of a gap in first and last exons of EgDGAT2_2 due to the different number of exons between EgDGAT2_1 and EgDGAT2_2. The number of exons varied between 5 and 9 exons as described by [25] although we found that the number was always 9 in *E. guineensis* and its relatives (Fig5B).

The difference between two EgDGAT3 with the closely related species is that date palm has a unique intron at the beginning of the sequence while one of the banana sequences (GSMUA_Achr1P25080_001) has one at the 3' end. For EgWS/DGAT, it is important to split the mapping alignment of EgWS/DGAT_1, EgWS/DGAT_2 and EgWS/DGAT_3 due to separation of these genes into two major clades. Interestingly, *E. guineensis* EgWS/DGAT_1 has two very long introns, i.e. intron one (2844 bp) and intron five (1497 bp). For EgWS/DGAT_2, we mapped the sequence to an EG5.1 chromosome, as we identified it as full length in the latest *E. guineensis* gene model (unpublished data). In Fig. 5E, we show that PDK_30s1002811g004 is similar to EgWS/DGAT_2 with conservation of almost all exons and introns. The only difference is that the date palm sequences have an additional intron (Intron 5). More information about the number of intron and exon in DGAT family is shown in Supplementary S1 file.

Chromosomal location of DGAT genes in E. guineensis

The diploid *E. guineensis* genome is made up of 16 chromosome pairs which for the purpose of sequencing assemblies are referred to as pseudochromosomes. A total of eight *E. guineensis* DGAT genes were mapped onto the 16 pseudochromosomes of the EG5.1 assembly in order to identify the precise location of these genes. A schematic diagram of the gene positions is given in Fig. 6. More detail information about the exact locations of these genes is given in Supplementary S1 file. DGAT1 genes were present in three different chromosomes, of which EgDGAT1_1 and EgDGAT1_2 generated hits at three chromosomes (3, 6 and 7), while EgDGAT1_3 only hit to chromosomes 6 and 7 (EgChr 6 and 7). To further reveal which EgDGAT1 sequences were associated with which chromosomes, a mapping analysis was used to check the size of exons and introns between the three EgDGAT1 sequences on the three chromosomes. From this analysis it was concluded that DGAT1 has probably undergone several duplication events. In particular, for EgDGAT1_1 there is a duplication between EgChr3 and EgChr7 with scores of 2, 483 and 2, 095 respectively. Therefore, although the three EgDGAT1 genes are only present in three chromosomal loci, some of these loci probably contain two tandemly and segmentally duplicated gene sequences. In particular EgDGAT1_2 appears to be segmentally duplicated on EgChr6 and EgChr7 while EgDGAT1_3 is duplicated in EgChr6 and EgChr7. From the MyPalmviewer Gbrowse website (http://gbrowse.mpob.gov.my/fgb2/gbrowse/Eg5_1/), additional evidence showed these genes were mapped at the same location with Arabidopsis (AT2G19450.1) on chromosomes 3, 6 and 7. A summary of the size (bp) exon and intron is illustrated in Supplementary Fig. S6.

In the case of DGAT2, the two genes hit to two different chromosomes where EgDGAT2_1 is on chromosome 10 and EgDGAT2_2 is on chromosome 4. Therefore in this case there is a single gene at each locus. For DGAT3, EgDGAT3_1 was located on EgChr2 while EgDGAT3_2 hit to EG_Un_random5. Finally, EgWS/DGAT_1 is present on EgChr6 while EgWS/DGAT_2 and the likely pseudogene, EgWS/DGAT_3, are both located on EgChr12, possibly due to a former tandem duplication event.

Transcript quantification and differential expression analysis

Quantitative analysis of the expression of eight DGAT genes during the development of endosperm and mesocarp tissues is shown in Figs. 7A and B. It is noteworthy that the patterns of gene expression of the various DGAT isoforms are completely different in these two fruit tissues. In particular, the expression levels of all DGAT genes are much less variable in the maternal mesocarp tissue compared with the triploid endosperm tissue that makes up the bulk of the seed kernel in *E. guineensis*. The three genes that are most highly upregulated in the endosperm are EgDGAT1_1, EgDGAT3_1 and EgWS/DGAT_1. Interestingly in the endosperm tissue, the expression levels increase after 12 weeks after anthesis (WAA) for EgDGAT1_1, EgDGAT2_1, EgDGAT3_1 and EgWS/DGAT_1. EgDGAT1_2 showed higher expression in endosperm but lower expression in the mesocarp, while the pattern was opposite for EgDGAT3_2. Also, while EgDGAT1_3 was highly expressed in both tissues, the opposite was found for EgDGAT2_2. Up-regulation of EgDGAT1_1 in endosperm and mesocarp at time point 12WAA and 15WAA respectively suggested the involvement of this particular isoform during oil accumulation [5].

The important role that DGAT1 plays in plant TAG synthesis has been demonstrated by over-expression studies [21,23,24] and by mutations in the DGAT1 gene [19]. Structural features in DGAT1 that lead to increased activity and enhanced oil yields when expressed in crops have been reported recently by [54]. These studies have expanded the previous research by the same group and also observations of amino acid substitutions in maize DGAT by [94]. Within *E. guineensis* fruits its is the kernel located endosperm tissue that is mainly involved in the formation of medium chain fatty acids (i.e. lauric and myristic) [95,96]. It has been reported that expression levels of DGAT1 and DGAT2 are higher in seed compared to mesocarp tissues [42,97]. Our endosperm transcriptome data for EgDGAT2_1 showed a similar pattern to EgDGAT1_3. This trend is similar to that reported in castor bean [98]. Hence, at 15WAA the *E. guineensis* EgDGAT2_1 was more highly expressed compared to all three EgDGAT1 isoforms. DGAT2 was first reported by [99] and may have a different expression profile to DGAT1 in some plants where it has been associated with the synthesis of TAGs with a unusual fatty acid compositions, such as in castor bean or tung tree [79,98,100,101]. DGAT2 transcripts are also found at relatively high levels in olive [102] and palm fruits [56,58]. Not surprisingly, in view of their quite distinct amino acid sequences, DGAT1 and DGAT2 have been reported to have different substrate selectivities [79,103–105] and this may account for their preferential use in different plants where different fatty acids are stored in TAGs.

In Fig. 7C a heatmap analysis shows the expression pattern for the *E. guineensis* DGAT2, DGAT3 and WS/DGAT families within 22 transcriptome libraries. No EgDGAT1 profiles are shown in the heatmap because of insufficient transcriptome data coverage but the

sequences are partially mapped to the genome. The results in Fig 7C suggest that for EgDGAT2_1, the major expression occurs in Mesocarp_1 (10WAA Mesocarp (DxP)) tissue. The expression profile for both EgDGAT3 is active in almost all transcriptome libraries. EgDGAT3_1 was highly expressed in Early_fruit_1 (Tenera floret before anthesis (BA)), Kernel_2 (15WAA Kernel (DxP)) and Shoot_1 (Shoot apex of Normal DxP clone). Based on the profiles showed in Fig 7B, EgDGAT3_2 was more highly expressed in mesocarp than other DGATs. There is a possibility that EgDGAT3 has a similar function to EgDGAT1 and is maybe involved in the TAG biosynthesis. In addition, [106] showed that the expression pattern of DGAT3 in Arabidopsis was similar to that for DGAT1 during seed development, although DGAT3 expression was higher during late maturation. Even though we found the presence of two full-length WS/DGATgenes in *E. guineensis*, only EgWS/DGAT_1 expression was detected in endosperm and mesocarp. The expression in endosperm was detected after 12WAA. Furthermore, this multifunctional / diverse function protein was highly expressed in two very different tissues, namely Pollen_2 (Pollen Pisifera (Fertile)) and Root_2 (Dura root). Gene expression was also observed in inflorescence and shoot tissues, which correlates well with the profiles reported by [47].

In conclusion, we present an updated and more detailed account of the DGAT gene families in *E. guineensis* that is based on the most recent gene model [7] and includes the characterization of two isoforms in each of the four DGAT gene families. Hence a total of eight functional DGAT sequences was identified using a comparative genomics approach. We have also compared the gene structure of *E. guineensis* DGATs with those of the closely related species *M. acuminata*, *P. dactylifera*. The activity of oil palm DGAT gene families was measured in detail by transcriptome profiling in various key tissues and developmental stages. Trascriptonal analysis shows that the oil palm genome contains several copies of each of the four DGAT gene families which are subject to differential regulation in different tissues and developmental stages, not only during fruit development but also during other important physiological processes such as vegetative growth and flower development.

Acknowledgements

We thank the Director General of MPOB, Dr. Ahmad Kushairi Din, for his support and encouragement throughout the project. Support for this work was provided by a PhD scholarship from MPOB to R Rosli to do research at the University of South Wales, UK.

References

[1] J.L. Harwood, H.K. Woodfield, G. Chen, R.J. Weselake, Modification of Oil Crops to Produce Fatty Acids for Industrial Applications, in: M.U. Ahmad (Ed.), Fat. AcidsChemistry, Synth. Appl., Elsevier, 2017: pp. 187–236. doi:10.1016/B978-0-12-809521-8.00005-2.

[2] D.J. Murphy, The future of oil palm as a major global crop : opportunities and challenges, J. Oil Palm Res. 26 (2014) 1–24.

[3] A. Kushairi, R. Singh, M. Ong Abdullah, The Oil Palm Industry in Malaysia : Thriving With Transformative Technologies, J. Oil Palm Res. 29 (2017) 431–439. doi:10.21894/jopr.2017.00017.

[4] F.D. Gunstone, J.L. Harwood, A.J. Dijkstra, eds., The Lipid Handbook, 3rd ed, CRC Press Taylor & Francis Group, 2007.

[5] R. Sambanthamurthi, K. Sundram, Y. Tan, Chemistry and biochemistry of palm oil, Prog Lipid Res. 39 (2000). doi:10.1016/S0163-7827(00)00015-1.

[6] R. Singh, M. Ong-Abdullah, E.L. Low, M.A.A. Manaf, R. Rosli, R. Nookiah, L.C.-L. Ooi, S.-E. Ooi, K. Chan, M.A. Halim, N. Azizi, J. Nagappan, B. Bacher, N. Lakey, S.W. Smith, D. He, M. Hogan, M.A. Budiman, E.K. Lee, R. DeSalle, D. Kudrna, J.L. Goicoechea, R.A. Wing, R.K. Wilson, R.S. Fulton, J.M. Ordway, R.A. Martienssen, R. Sambanthamurthi, Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds., Nature. 500 (2013) 335–359. doi:10.1038/nature12309.

[7] K.-L. Chan, T. V. Tatarinova, R. Rosli, N. Amiruddin, N. Azizi, M.A.A. Halim, N.S.N.M. Sanusi, N. Jayanthi, P. Ponomarenko, M. Triska, V. Solovyev, M. Firdaus-Raih, R. Sambanthamurthi, D. Murphy, E.-T.L. Low, Evidence-based gene models for structural and functional annotations of the oil palm genome, Biol. Direct. (2017). doi:10.1186/s13062-017-0191-4.

[8] D.J. Murphy, The dynamic roles of intracellular lipid droplets: from archaea to mammals, 249 (2012) 541–585. doi:10.1007/s00709-011-0329-7.

[9] P.D. Bates, S. Stymne, J. Ohlrogge, Biochemical pathways in seed oil synthesis, Curr. Opin. Plant Biol. 16 (2013) 358–364. doi:10.1016/j.pbi.2013.02.015.

[10] G. Chen, H.K. Woodfield, X. Pan, J.L. Harwood, R.J. Weselake, Acyl-Trafficking during Plant Oil Accumulation, Lipids. 50 (2015) 1057–1068. doi:10.1007/s11745-015-4069-x.

[11] R.J. Weselake, D.C. Taylor, M.H. Rahman, S. Shah, A. Laroche, P.B.E. McVetty, J.L. Harwood, Increasing the flow of carbon into seed oil, Biotechnol. Adv. 27 (2009) 866–878. doi:10.1016/j.biotechadv.2009.07.001.

[12] R.J. Weselake, Storage Lipids, in: D.J. Murphy (Ed.), Plant Lipids Biol. Util. Manip., Blackwell Publishing,Oxford., 2005: pp. 162–225.

[13] P.D. Bates, Understanding the control of acyl flux through the lipid metabolic network of plant oil biosynthesis, Biochim. Biophys. Acta - Mol. Cell Biol. Lipids. (2016). doi:10.1016/j.bbalip.2016.03.021.

[14] E.P. Kennedy, Biosynthesis of complex lipids., Fed. Proc. 20 (1961) 934–940.

[15] H.J. Perry, J.L. Harwood, Radiolabelling studies of acyl lipids in developing seeds of Brassica napus: Use of [1-14C]acetate precursor, Phytochemistry. 33 (1993) 329–333. doi:10.1016/0031-9422(93)85512-P.

[16] H.J. Perry, J.L. Harwood, Changes in the lipid content of developing seeds of Brassica napus, Phytochemistry. 32 (1993) 1411–1415. doi:10.1016/0031-9422(93)85148-K.

[17] H.J. Perry, R. Bligny, E. Gout, J.L. Harwood, Changes in Kennedy pathway intermediates associated with increased triacylglycerol synthesis in oil-seed rape, Phytochemistry. 52 (1999) 799–804. doi:10.1016/S0031-9422(99)00294-0.

[18] V. Katavic, D.W. Reed, D.C. Taylor, E.M. Giblin, D.L. Barton, J. Zou, S.L. Mackenzie, P.S. Covello, L. Kunst, Alteration of seed fatty acid composition by an ethyl methanesulfonate-induced mutation in Arabidopsis thaliana affecting diacylglycerol acyltransferase activity., Plant Physiol. 108 (1995) 399–409. doi:10.1104/pp.108.1.399.

[19] J. Zou, Y. Wei, C. Jako, A. Kumar, G. Selvaraj, D.C. Taylor, The Arabidopsis thaliana TAG1 mutant has a mutation in a diacylglycerol acyltransferase gene, Plant J. 19 (1999) 645–653. doi:10.1046/j.1365-313X.1999.00555.x.

[20] J.M. Routaboul, C. Benning, N. Bechtold, M. Caboche, L. Lepiniec, The TAG1 locus of *Arabidopsis* encodes for a diacylglycerol acyltransferase, Plant Physiol. Biochem. 37 (1999) 831–840. doi:10.1016/S0981-9428(99)00115-1.

[21] C. Jako, A. Kumar, Y. Wei, J. Zou, D.L. Barton, E.M. Giblin, P.S. Covello, D.C. Taylor, Seed-specific over-expression of an Arabidopsis cDNA encoding a diacylglycerol acyltransferase enhances seed oil content and seed weight., Plant Physiol. 126 (2001) 861–74. doi:10.1104/PP.126.2.861.

[22] M. Tang, I.A. Guschina, P. O'Hara, A.R. Slabas, P.A. Quant, T. Fawcett, J.L. Harwood, Metabolic control analysis of developing oilseed rape (Brassica napus cv Westar) embryos shows that lipid assembly exerts significant control over oil accumulation, New Phytol. 196 (2012) 414–426. doi:10.1111/j.1469-8137.2012.04262.x.

[23] R.J. Weselake, S. Shah, M. Tang, P.A. Quant, C.L. Snyder, T.L. Furukawa-Stoffer, W. Zhu, D.C. Taylor, J. Zou, A. Kumar, L. Hall, A. Laroche, G. Rakow, P. Raney, M.M. Moloney, J.L. Harwood, Metabolic control analysis is helpful for informed genetic manipulation of oilseed rape (Brassica napus) to increase seed oil content, J. Exp. Bot. 59 (2008) 3543–3549. doi:10.1093/jxb/ern206.

[24] D.C. Taylor, Y. Zhang, A. Kumar, T. Francis, E.M. Giblin, D.L. Barton, J.R. Ferrie, A. Laroche, S. Shah, W. Zhu, Molecular modification of triacylglycerol accumulation by over-expression of DGAT1 to produce canola with increased seed oil content under field conditions, Botany. 87 (2009) 533–543. doi:10.1139/B08-101.

[25] A.C. Turchetto-Zolet, A.P. Christoff, F.R. Kulcheski, G. Loss-Morais, R. Margis, M. Margis-Pinheiro, Diversity and evolution of plant diacylglycerol acyltransferase (DGATs) unveiled by phylogenetic, gene structure and expression analyses, Genet. Mol. Biol. (2016). doi:10.1590/1678-4685-GMB-2016-0024.

[26] J. Tzen, Y. Cao, P. Laurent, C. Ratnayake, A. Huang, Lipids, proteins, and structure of seed oil bodies from diverse species., Plant Physiol. 101 (1993) 267–276. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=158673&tool=pmcentrez&rendertype=abstract.

[27] A.R. Slabas, C.M. Sidebottom, A. Hellyer, R.M.J. Kessell, M.P. Tombs, Induction, purification and characterization of NADH-specific enoyl acyl carrier protein reductase from developing seeds of oil seed rape (Brassica napus), Biochim. Biophys. Acta - Lipids Lipid Metab. 877 (1986) 271–280. doi:10.1016/0005-2760(86)90304-8.

[28] A.R. Slabas, J. Harding, A. Hellyer, P. Roberts, H.E. Bambridge, Induction, purification and characterization of acyl carrier protein from developing seeds of oil seed rape (Brassica napus), Biochim. Biophys. Acta - Lipids Lipid Metab. 921 (1987) 50–59. doi:https://doi.org/10.1016/0005-2760(87)90169-X.

[29] R.W. Mackintosh, D.G. Hardie, A.R. Slabas, A new assay procedure to study the induction of Beta-ketoacyl-ACP synthase I and II, and the complete purification of Beta-ketoacyl-ACP synthase I from developing seeds of oilseed Rape (Brassica napus), Biochim. Biophys. Acta. 1002 (1989) 114–124.

[30] U.S. Ramli, D.S. Baker, P. a Quant, J.L. Harwood, Control analysis of lipid

biosynthesis in tissue cultures from oil crops shows that flux control is shared between fatty acid synthesis and lipid assembly., Biochem. J. 364 (2002) 393–401. doi:10.1042/BJ20010203.

[31] U.S. Ramli, D.S. Baker, P.A. Quant, J.L. Harwood, Control mechanisms operating for lipid biosynthesis differ in oil-palm (Elaeis guineensis Jacq.) and olive (Olea europaea L.) callus cultures, Biochem. J. 364 (2002) 385–391. doi:10.1042/BJ20010202.

[32] U.S. Ramli, J.J. Salas, P.A. Quant, J.L. Harwood, Metabolic control analysis reveals an important role for diacylglycerol acyltransferase in olive but not in oil palm lipid accumulation, FEBS J. 272 (2005) 5764–5770. doi:10.1111/j.1742-4658.2005.04964.x.

[33] U.S. Ramli, J.J. Salas, P.A. Quant, J.L. Harwood, Use of metabolic control analysis to give quantitative information on control of lipid biosynthesis in the important oil crop, Elaeis guineensis (oilpalm)., New Phytol. 184 (2009) 330–9. doi:10.1111/j.1469-8137.2009.02962.x.

[34] P. Bouvier-Navé, P. Benveniste, P. Oelkers, S.L. Sturley, H. Schaller, Expression in yeast and tobacco of plant cDNAs encoding acyl CoA:diacylglycerol acyltransferase, Eur. J. Biochem. 267 (2000) 85–96. doi:10.1046/j.1432-1327.2000.00961.x.

[35] T. Vanhercke, A. El Tahchy, P. Shrestha, X.R. Zhou, S.P. Singh, J.R. Petrie, Synergistic effect of WRI1 and DGAT1 coexpression on triacylglycerol biosynthesis in plants, FEBS Lett. (2013). doi:10.1016/j.febslet.2012.12.018.

[36] K. Meyer, A.J. Kinney, Biosynthesis and biotechnology of seed lipids including sterols, carotenoids and tocochromanols, in: H. Wada, N. Murata (Eds.), Adv. Photosynth. Respir. (Lipids Photosynth., Volume 30, Springer, New York, 2009: pp. 407–444.

[37] K. Meyer, K.L. Stecca, DGAT genes for increased seed storage lipid production and altered fatty acid profiles in oilseed plants., 2017.

[38] V. Andrianov, N. Borisjuk, N. Pogrebnyak, A. Brinker, J. Dixon, S. Spitsin, J. Flynn, P. Matyszczuk, K. Andryszak, M. Laurelli, M. Golovkin, H. Koprowski, Tobacco as a production platform for biofuel : overexpression of Arabidopsis DGAT and LEC2 genes increases accumulation and shifts the composition of lipids in green biomass, (2010) 277–287. doi:10.1111/j.1467-7652.2009.00458.x.

[39] T. Vanhercke, U.K. Divi, A. El Tahchy, Q. Liu, M. Mitchell, M.C. Taylor, P.J. Eastmond, F. Bryant, A. Mechanicos, C. Blundell, Y. Zhi, S. Belide, P. Shrestha, X.R. Zhou, J.P. Ral, R.G. White, A. Green, S.P. Singh, J.R. Petrie, Step changes in leaf oil accumulation via iterative metabolic engineering, Metab. Eng. 39 (2017) 237–246. doi:10.1016/j.ymben.2016.12.007.

[40] S.B. Weiss, E.P. Kennedy, J.Y. Kiyasu, The enzymatic of triglycerides, J. Biol. Chem. 235 (1960) 40–44.

[41] A. Cagliari, R. Margis, F. Dos Santos Maraschin, A.C. Turchetto-Zolet, G. Loss, M. Margis-Pinheiro, Biosynthesis of triacylglycerols (TAGs) in plants and algae, Int. J. Plant Biol. 2 (2011) 40–52. doi:10.4081/pb.2011.e10.

[42] Q. Liu, R.M.P. Siloto, R. Lehner, S.J. Stone, R.J. Weselake, Acyl-CoA:diacylglycerol acyltransferase: Molecular biology, biochemistry and biotechnology, Prog. Lipid Res. 51 (2012) 350–377. doi:10.1016/j.plipres.2012.06.001.

[43] A.C. Turchetto-Zolet, F.S. Maraschin, G.L. De Morais, A. Cagliari, C.M.B. Andrade, M. Margis-Pinheiro, R. Margis, Evolutionary view of acyl-CoA diacylglycerol acyltransferase (DGAT), a key enzyme in neutral lipid biosynthesis, BMC Evol. Biol. 11 (2011) 263. doi:10.1186/1471-2148-11-263.

[44] S. Saha, B. Enugutti, S. Rajakumari, R. Rajasekharan, Cytosolic triacylglycerol biosynthetic pathway in oilseeds. Molecular cloning and expression of peanut

cytosolic diacylglycerol acyltransferase., Plant Physiol. 141 (2006) 1533–1543. doi:10.1104/pp.106.082198.

[45] M.L. Hernández, L. Whitehead, Z. He, V. Gazda, A. Gilday, E. Kozhevnikova, F.E. Vaistij, T.R. Larson, I. a Graham, A cytosolic acyltransferase contributes to triacylglycerol synthesis in sucrose-rescued Arabidopsis seed oil catabolism mutants., Plant Physiol. 160 (2012) 215–25. doi:10.1104/pp.112.201541.

[46] R. Kalscheuer, A. Steinbüchel, A novel bifunctional wax ester synthase/acyl-CoA: diacylglycerol acyltransferase mediates wax ester and triacylglycerol biosynthesis in Acinetobacter calcoaceticus ADP1, J. Biol. Chem. 278 (2003) 8075–8082. doi:10.1074/jbc.M210533200.

[47] F. Li, X. Wu, P. Lam, D. Bird, H. Zheng, L. Samuels, R. Jetter, L. Kunst, Identification of the wax ester synthase/acyl-coenzyme A: diacylglycerol Acyltransferase WSD1 required for stem wax ester biosynthesis in Arabidopsis, Plant Physiol. 148 (2008) 97–107. doi:10.1104/pp.108.123471.

[48] T. Stöveken, R. Kalscheuer, U. Malkus, R. Reichelt, A. Steinbüchel, The Wax Ester Synthase / Acyl Coenzyme A : Diacylglycerol Acyltransferase from Acinetobacter sp . Strain ADP1 : Characterization of a Novel Type of Acyltransferase, J. Bacteriol. 187 (2005) 1369–1376. doi:10.1128/JB.187.4.1369.

[49] J.A. Villa, M. Cabezas, F. de la Cruz, G. Moncalián, Use of limited proteolysis and mutagenesis to identify folding domains and sequence motifs critical for wax ester synthase/acyl coenzyme A: diacylglycerol acyltransferase activity, Appl. Environ. Microbiol. 80 (2014) 1132–1141. doi:10.1128/AEM.03433-13.

[50] T. Tomiyama, K. Kurihara, T. Ogawa, T. Maruta, T. Ogawa, D. Ohta, Y. Sawa, T. Ishikawa, Wax ester synthase/diacylglycerol acyltransferase isoenzymes play a pivotal role in wax ester biosynthesis in Euglena gracilis, Sci. Rep. 7 (2017) 1–13. doi:10.1038/s41598-017-14077-6.

[51] J.E. Chen, A.G. Smith, A look at diacylglycerol acyltransferases (DGATs) in algae, J. Biotechnol. 162 (2012) 28–39. doi:10.1016/j.jbiotec.2012.05.009.

[52] T.-Y. Seng, E. Ritter, S.H. Mohamed Saad, L.-J. Leao, R.S. Harminder Singh, F. Qamaruz Zaman, S.-G. Tan, S.S.R. Syed Alwee, V. Rao, QTLs for oil yield components in an elite oil palm (Elaeis guineensis) cross, Euphytica. 212 (2016) 399–425. doi:10.1007/s10681-016-1771-6.

[53] C. Montoya, B. Cochard, A. Flori, D. Cros, R. Lopes, T. Cuellar, S. Espeout, I. Syaputra, P. Villeneuve, M. Pina, E. Ritter, T. Leroy, N. Billotte, Genetic architecture of palm oil fatty acid composition in cultivated oil palm (Elaeis guineensis Jacq.) compared to its wild relative E. oleifera (H.B.K) Cortés, PLoS One. 9 (2014). doi:10.1371/journal.pone.0095412.

[54] Y. Xu, G. Chen, M.S. Greer, K.M.P. Caldo, G. Ramakrishnan, S. Shah, L. Wu, M.J. Lemieux, J. Ozga, R.J. Weselake, Multiple mechanisms contribute to increased neutral lipid accumulation in yeast producing recombinant variants of plant diacylglycerol acyltransferase 1, J. Biol. Chem. 292 (2017) 17819–17831. doi:10.1074/jbc.M117.811489.

[55] F. Morcillo, D. Cros, N. Billotte, G.F. Ngando-Ebongue, H. Domonhédo, M. Pizot, T. Cuéllar, S. Espéout, R. Dhouib, F. Bourgis, S. Claverol, T.J. Tranbarger, B. Nouy, V. Arondel, Improving palm oil quality through identification and mapping of the lipase gene causing oil deterioration, Nat. Commun. 4 (2013) 1–8. doi:10.1038/ncomms3160.

[56] F. Bourgis, A. Kilaru, X. Cao, G.-F. Ngando-Ebongue, N. Drira, J.B. Ohlrogge, V. Arondel, Comparative transcriptome and metabolite analysis of oil palm and date palm mesocarp that differ dramatically in carbon partitioning., Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 12527–12532. doi:10.1073/pnas.1106502108.

[57] Y. Jin, Y. Yuan, L. Gao, R. Sun, L. Chen, D. Li, Y. Zheng, Characterization and functional analysis of a type 2 Diacylglycerol acyltransferase (DGAT2) gene from oil palm (Elaeis guineensis Jacq.) mesocarp in Saccharomyces cerevisiae and transgenic Arabidopsis thaliana, Front. Plant Sci. 8 (2017) 1–10. doi:10.3389/fpls.2017.01791.

[58] T.J. Tranbarger, S. Dussert, T. Joet, X. Argout, M. Summo, a. Champion, D. Cros, a. Omore, B. Nouy, F. Morcillo, Regulatory Mechanisms Underlying Oil Palm Fruit Mesocarp Maturation, Ripening, and Functional Specialization in Lipid and Carotenoid Metabolism, Plant Physiol. 156 (2011) 564–584. doi:10.1104/pp.111.175141.

[59] L. Li, C.J. Stoeckert, D.S. Roos, OrthoMCL: Identification of ortholog groups for eukaryotic genomes, Genome Res. 13 (2003) 2178–2189. doi:10.1101/gr.1224503.

[60] S.R. Eddy, Accelerated profile HMM searches, PLoS Comput. Biol. 7 (2011). doi:10.1371/journal.pcbi.1002195.

[61] S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipmann, Gapped BLAST and PSI- BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402. doi:10.1093/nar/25.17.3389.

[62] T.L. Bailey, M. Boden, F.A. Buske, M. Frith, C.E. Grant, L. Clementi, J. Ren, W.W. Li, W.S. Noble, MEME Suite: Tools for motif discovery and searching, Nucleic Acids Res. 37 (2009). doi:10.1093/nar/gkp335.

[63] M. Kearse, R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, A. Drummond, Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data, Bioinformatics. 28 (2012) 1647–1649. doi:10.1093/bioinformatics/bts199.

[64] R.C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792–1797. doi:10.1093/nar/gkh340.

[65] T.A. Hall, BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, Nucleic Acids Symp. Ser. 41 (1999) 95–98. doi:citeulike-article-id:691774.

[66] S. Kumar, G. Stecher, K. Tamura, MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets, Mol. Biol. Evol. 33 (2016) 1870–1874. doi:10.1093/molbev/msw054.

[67] N. Saitou, M. Nei, N.M. Saitou N, The neighbour-joining method: A new method for reconstructing phylogenetic trees, Mol. Biol. Evol. 4 (1987) 406–425. doi:citeulike-article-id:93683.

[68] E. Zuckerkandl, L. Pauling, Evolutionary divergence and convergence in proteins, New York Acad. Press Acad. Press. Evol. Genes Proteins, Ed. V. Bryson H. J. Vogel. (1965) 97–166. doi:10.1209/epl/i1998-00224-x.

[69] D.T. Jones, W.R. Taylor, J.M. Thornton, The rapid generation of mutation data matricies from protein sequences, Comput. Appl. Biosci. 8 (1992) 275–282. doi:10.1093/bioinformatics/8.3.275.

[70] G.S.C. Slater, E. Birney, Automated generation of heuristics for biological sequence comparison, BMC Bioinformatics. 6 (2005). doi:10.1186/1471-2105-6-31.

[71] F. Odronitz, H. Pillmann, O. Keller, S. Waack, M. Kollmar, WebScipio: An online tool for the determination of gene structures using protein sequences, 13 (2008) 1–13. doi:10.1186/1471-2164-9-422.

[72] K. Hatje, O. Keller, B. Hammesfahr, H. Pillmann, S. Waack, M. Kollmar, Cross-species protein sequence and gene structure prediction with fine-tuned Webscipio 2.0 and Scipio, BMC Res. Notes. 4 (2011) 265. doi:10.1186/1756-0500-4-265.

[73] B. Hammesfahr, F. Odronitz, S. Mühlhausen, S. Waack, M. Kollmar, GenePainter: A fast tool for aligning gene structures of eukaryotic protein families, visualizing the alignments and mapping gene structures onto protein structures, BMC Bioinformatics. 14 (2013). doi:10.1186/1471-2105-14-77.

[74] C. Trapnell, A. Roberts, L. Goff, G. Pertea, D. Kim, D.R. Kelley, H. Pimentel, S.L. Salzberg, J.L. Rinn, L. Pachter, Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks., Nat. Protoc. 7 (2012) 562–78. doi:10.1038/nprot.2012.016.

[75] J. Bousquet, S.H. Strauss, A.H. Doerksen, R.A. Price, Extensive variation in evolutionary rate of rbcL gene sequences among seed plants., Proc. Natl. Acad. Sci. 89 (1992) 7844–7848. doi:10.1073/pnas.89.16.7844.

[76] B.S. Gaut, B.R. Morton, B.C. McCaig, M.T. Clegg, Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene Adh parallel rate differences at the plastid gene rbcL., Proc. Natl. Acad. Sci. 93 (1996) 10274–10279. http://www.pnas.org/content/93/19/10274 (accessed April 4, 2018).

[77] C.F. Barrett, J.I. Davis, J. Leebens-Mack, J.G. Conran, D.W. Stevenson, Plastid genomes and deep relationships among the commelinid monocot angiosperms, Cladistics. 29 (2013) 65–87. doi:10.1111/j.1096-0031.2012.00418.x.

[78] C.F. Barrett, W.J. Baker, J.R. Comer, J.G. Conran, S.C. Lahmeyer, J.H. Leebens-Mack, J. Li, G.S. Lim, D.R. Mayfield-Jones, L. Perez, J. Medina, J.C. Pires, C. Santos, D. Wm. Stevenson, W.B. Zomlefer, J.I. Davis, Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots, New Phytol. 209 (2016) 855–870. doi:10.1111/nph.13617.

[79] J.M. Shockey, Tung tree DGAT1 and DGAT2 have nonredundant functions in triacylglycerol biosynthesis and are localized to different subdomains of the endoplasmic reticulum, Plant Cell Online. 18 (2006) 2294–2313. doi:10.1105/tpc.106.043695.

[80] S. Cases, S.J. Stone, P. Zhou, E. Yen, B. Tow, K.D. Lardizabal, T. Voelker, R. V. Farese, Cloning of DGAT2, a Second Mammalian Diacylglycerol Acyltransferase, and Related Family Members, J. Biol. Chem. 276 (2001) 38870–38876. doi:10.1074/jbc.M106219200.

[81] S. Dussert, C. Guerin, M. Andersson, T. Joet, T.J. Tranbarger, M. Pizot, G. Sarah, A. Omore, T. Durand-Gasselin, F. Morcillo, Comparative Transcriptome Analysis of Three Oil Palm Fruit and Seed Tissues That Differ in Oil Content and Fatty Acid Composition, Plant Physiol. 162 (2013) 1337–1358. doi:10.1104/pp.113.220525.

[82] M. Frentzen, F. Wolter, Molecular biology of acyltransferases involved in glycerolipid synthesis., Plant Lipid Biosynthesisfundamentals Agric. Appl. Exp. Biol. Semin. Ser. 67 (1998) 247–272.

[83] C.J.A. Sigrist, E. De Castro, L. Cerutti, B.A. Cuche, N. Hulo, A. Bridge, L. Bougueleret, I. Xenarios, New and continuing developments at PROSITE, Nucleic Acids Res. 41 (2013) 344–347. doi:10.1093/nar/gks1067.

[84] J.L.S. Lopes, T.M. Nobre, E.M. Cilli, L.M. Beltramini, A.P.U. Araújo, B.A. Wallace, Deconstructing the DGAT1 enzyme: Binding sites and substrate interactions, Biochim. Biophys. Acta - Biomembr. 1838 (2014) 3145–3152. doi:10.1016/j.bbamem.2014.08.017.

[85] P.J. McFie, S.L. Stone, S.L. Banman, S.J. Stone, Topological orientation of acyl-CoA:Diacylglycerol acyltransferase-1 (DGAT1) and identification of a putative active site histidine and the role of the N terminus in dimer/tetramer formation, J. Biol. Chem. 285 (2010) 37377–37387. doi:10.1074/jbc.M110.163691.

[86] C.-L.E. Yen, S.J. Stone, S. Koliwad, C. Harris, R. V. Farese, *Thematic Review Series:*

Glycerolipids. DGAT enzymes and triacylglycerol biosynthesis, J. Lipid Res. 49 (2008) 2283–2301. doi:10.1194/jlr.R800018-JLR200.

[87] R.J. Weselake, M. Madhavji, S.J. Szarka, N.A. Patterson, W.B. Wiehler, C.L. Nykiforuk, T.L. Burton, P.S. Boora, S.C. Mosimann, N.A. Foroud, B.J. Thibault, M.M. Moloney, A. Laroche, T.L. Furukawa-Stoffer, Acyl-CoA-binding and self-associating properties of a recombinant 13.3 kDa N-terminal fragment of diacylglycerol acyltransferase-1 from oilseed rape, BMC Biochem. 7 (2006). doi:10.1186/1471-2091-7-24.

[88] Y. Jin, P.J. McFie, S.L. Banman, C. Brandt, S.J. Stone, Diacylglycerol acyltransferase-2 (DGAT2) and monoacylglycerol acyltransferase-2 (MGAT2) interact to promote triacylglycerol synthesis, J. Biol. Chem. 289 (2014) 28237–28248. doi:10.1074/jbc.M114.571190.

[89] K. Hofmann, A superfamily of membrane-bound O-acyltransferases with implications for Wnt signaling, Trends Biochem. Sci. 25 (2000) 111–112. doi:10.1016/S0968-0004(99)01539-X.

[90] S.J. Stone, M.C. Levin, R. V. Farese, Membrane topology and identification of key functional amino acid residues of murine Acyl-CoA:diacylglycerol acyltransferase-2, J. Biol. Chem. 281 (2006) 40273–40282. doi:10.1074/jbc.M607986200.

[91] M.R. Owen, C.C. Corstorphine, V.A. Zammit, Overt and latent activities of diacylglycerol acytransferase in rat liver microsomes: possible roles in very-low-density lipoprotein triacylglycerol secretion., Biochem. J. 323 (1997) 17–21. doi:10.1042/bj3230017.

[92] K.A.H. Abo-Hashema, M.H. Cake, G.W. Power, D. Clarke, Evidence for triacylglycerol synthesis in the lumen of microsomes via a lipolysis-esterification pathway involving carnitine acyltransferases, J. Biol. Chem. 274 (1999) 35577–35582. doi:10.1074/jbc.274.50.35577.

[93] I.J. Waterman, N.T. Price, V.A. Zammit, Distinct ontogenic patterns of overt and latent DGAT activities of rat liver microsomes, J. Lipid Res. 43 (2002) 1555–1562. doi:10.1194/jlr.M200051-JLR200.

[94] P. Zheng, W.B. Allen, K. Roesler, M.E. Williams, S. Zhang, J. Li, K. Glassman, J. Ranch, D. Nubel, W. Solawetz, D. Bhattramakki, V. Llaca, S. Deschamps, G.Y. Zhong, M.C. Tarczynski, B. Shen, A phenylalanine in DGAT is a key determinant of oil content and composition in maize, Nat. Genet. 40 (2008) 367–372. doi:10.1038/ng.85.

[95] S.-Y. Kok, M. Ong-Abdullah, G. Cheng-Lian EE, P. Namasivayam, A histological study of oil palm (Elaeis guineensis) endosperm during seed development, J. Oil Palm Res. 27 (2015) 107–112.

[96] L. Aymé, P. Jolivet, J.M. Nicaud, T. Chardot, Molecular characterization of the elaeis guineensis medium-chain fatty acid diacylglycerol acyltransferase DGAT1-1 by heterologous expression in yarrowia lipolytica, PLoS One. (2015). doi:10.1371/journal.pone.0143113.

[97] R. Li, K. Yu, D.F. Hildebrand, DGAT1, DGAT2 and PDAT expression in seeds and other tissues of epoxy and hydroxy fatty acid accumulating plants, Lipids. 45 (2010) 145–157. doi:10.1007/s11745-010-3385-4.

[98] J.T.M. Kroon, W. Wei, W.J. Simon, A.R. Slabas, Identification and functional expression of a type 2 acyl-CoA:diacylglycerol acyltransferase (DGAT2) in developing castor bean seeds which has high homology to the major triglyceride biosynthetic enzyme of fungi and animals, Phytochemistry. (2006). doi:10.1016/j.phytochem.2006.09.020.

[99] K.D. Lardizabal, J.T. Mai, N.W. Wagner, A. Wyrick, T. Voelker, D.J. Hawkins,

DGAT2 is a new diacylglycerol acyltransferase gene family. Purification, cloning, and expression in insect cells of two polypeptides from Mortierella ramanniana with diacylglycerol acyltransferase activity, J. Biol. Chem. 276 (2001) 38862–38869. doi:10.1074/jbc.M106168200.

[100] J. Burgal, J. Shockey, C. Lu, J. Dyer, T. Larson, I. Graham, J. Browse, Metabolic engineering of hydroxy fatty acid production in plants: RcDGAT2 drives dramatic increases in ricinoleate levels in seed oil, Plant Biotechnol. J. 6 (2008) 819–831. doi:10.1111/j.1467-7652.2008.00361.x.

[101] F. Chen, A.J. Mackey, J.K. Vermunt, D.S. Roos, Assessing performance of orthology detection strategies applied to eukaryotic genomes, PLoS One. 2 (2007). doi:10.1371/journal.pone.0000383.

[102] F. Alagna, N. D'Agostino, L. Torchia, M. Servili, R. Rao, M. Pietrella, G. Giuliano, M.L. Chiusano, L. Baldoni, G. Perrotta, Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development, BMC Genomics. 10 (2009) 399. doi:10.1186/1471-2164-10-399.

[103] S.C. Lung, R.J. Weselake, Diacylglycerol acyltransferase: A key mediator of plant triacylglycerol synthesis, Lipids. 41 (2006) 1073–1088. doi:10.1007/s11745-006-5057-y.

[104] X.R. Zhou, P. Shrestha, F. Yin, J.R. Petrie, S.P. Singh, AtDGAT2 is a functional acyl-CoA:diacylglycerol acyltransferase and displays different acyl-CoA substrate preferences than AtDGAT1, FEBS Lett. 587 (2013) 2371–2376. doi:10.1016/j.febslet.2013.06.003.

[105] L. Aymé, S. Baud, B. Dubreucq, F. Joffre, T. Chardot, Function and localization of the Arabidopsis thaliana diacylglycerol acyltransferase DGAT2 expressed in yeast, PLoS One. 9 (2014) 1–9. doi:10.1371/journal.pone.0092237.

[106] F.Y. Peng, R.J. Weselake, Gene coexpression clusters and putative regulatory elements underlying seed storage reserve accumulation in Arabidopsis, BMC Genomics. 12 (2011). doi:10.1186/1471-2164-12-286.

Fig 1. Phylogenetic tree and motif analysis of three *E. guineensis* DGAT1 sequences with orthologs from 12 species. (A) DGAT1 phylogenetic tree (Different coloured boxes are used to distinguish between monocots (orange box) and dicots (green box)) (B) Conserved motifs distribution identified using MEME (Represented by coloured boxes). Species abbreviations: Ath: *A. thaliana*, Bdi: *B. distachyon*, Bna: *B. napus*, Pdi: *P. dactylifera*, Eo*: E. oleifera*, Gh: *G. hirsutum*, Gm: *G. max*, Ha: *H. annuus*, Mac: *M. acuminata*, Eg: *E. guineensis*, Osa: *O. sativa*, Sbi: *S. bicolor*, Zm: *Z. mays*.

Fig 2. Phylogenetic tree and motif analysis of two *E. guineensis* DGAT2 sequences with orthologs from 12 species (A) DGAT2 phylogenetic tree (Different coloured boxes are used to distinguish between monocots (orange box) and dicots (green box)), (B) Conserved motifs distribution identified using MEME (Represented by coloured boxes). Species abbreviations: Ath: *A. thaliana*, Bdi: *B. distachyon*, Bna: *B. napus*, Pdi: *P. dactylifera*, Eo*: E. oleifera*, Gh: *G. hirsutum*, Gm: *G. max*, Ha: *H. annuus*, Mac: *M. acuminata*, Eg: *E. guineensis*, Osa: *O. sativa*, Sbi: *S. bicolor*, Zm: *Z. mays*.

Fig 3. Phylogenetic tree and motif analysis of two *E. guineensis* DGAT3 sequences with orthologs from 12 species. (A) DGAT3 phylogenetic tree (Different coloured boxes are used to distinguish between monocots (orange box) and dicots (green box)). (B) Conserved motifs distribution identified using MEME (Represented by coloured boxes). Species abbreviations: Ath: *A. thaliana*, Bdi: *B. distachyon*, Bna: *B. napus*, Pdi: *P. dactylifera*, Eo*: E. oleifera*, Gh: *G. hirsutum*, Gm: *G. max*, Ha: *H. annuus*, Mac: *M. acuminata*, Eg: *E. guineensis*, Osa: *O. sativa*, Sbi: *S. bicolor*, Zm: *Z. mays*.

Fig 4. Phylogenetic tree and motif analysis of *E. guineensis* WS/DGAT sequences with orthologs from 12 species. (A) WS/DGAT phylogenetic tree (Different coloured boxes are used to distinguish between monocots (orange box) and dicots (green box)). (B) Conserved motifs distribution identified using MEME (Represented by coloured boxes). Species abbreviations: Ath: *A. thaliana*, Bdi: *B. distachyon*, Bna: *B. napus*, Pdi: *P. dactylifera*, Eo*: E. oleifera*, Gh: *G. hirsutum*, Gm: *G. max*, Ha: *H. annuus*, Mac: *M. acuminata*, Eg: *E. guineensis*, Osa: *O. sativa*, Sbi: *S. bicolor*, Zm: *Z. mays*.

Fig 5 Structural alignment of DGAT genes between *E. guineensis* and the two closely related species, banana and date palm. (A) DGAT1, (B) DGAT2, (C) DGAT3, (D & E) WS/DGAT. Introns are shown as purple boxes while orange boxes indicate exons.

Fig 6 Schematic view of DGAT gene locations on the 16 pseudochromosomes of *E. guineensis*.

Fig 7 Expression profile of the DGAT gene family in different *E. guineensis* tissue types. (A) DGAT expression profile in endosperm and (B) mesocarp for different points in weeks after anthesis (WAA). (C) Heatmap of expressed in 22 *E. guineensis* transcriptome libraries. Tissue library: Early_fruit_1 (Tenera floret before anthesis (BA)), Early_fruit_2 (Tenera

floret/fruit after anthesis (AA)), Early_fruit_3 (Dura floret before anthesis (BA)), Early_fruit_4 (Dura floret/fruit after anthesis (AA)), Pollen_1 (Pollen Pisifera (Sterile)), Pollen_2 (Pollen Pisifera (Fertile)), Inflorescence_1 (Female flower of Normal DxP clone (2.5cm)), Inflorescence_2 (Female flower of abnormal DxP clone (2cm)), Inflorescence_3 (Female flower of Normal DxP clone (20cm)), Inflorescence_4 (Female flower of abnormal DxP clone (19cm)), Kernel_1 (10WAA Kernel (DxP)), Kernel_2 (15WAA Kernel (DxP)), Mesocarp_1 (10WAA Mesocarp (DxP)), Mesocarp_2 (15WAA Mesocarp (DxP)), Leaf_1 (Spear Leaf Dura), Leaf_2 (Spear Leaf Pisifera), Leaf_3 (F17 Leaf Pisifera), Root_1 (Root whitish roots from seedling palms (DxP)), Root_2 (Dura root), Root_3 (Pisifera root), Shoot_1 (Shoot apex of Normal DxP clone), Shoot_2 (Shoot apex of abnormal DxP clone)